An empirical analysis of dockless bike-sharing utilization and its explanatory factors: Case study from Shanghai, China

Downloaded from: https://research.chalmers.se, 2023-05-04 19:51 UTC

Citation for the original published paper (version of record):
Li, A., Zhao, P., Huang, Y. et al (2020). An empirical analysis of dockless bike-sharing utilization and its explanatory factors: Case
study from Shanghai, China. Journal of Transport Geography, 88.
http://dx.doi.org/10.1016/j.jtrangeo.2020.102828

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology.
It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004.
research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Journal of Transport Geography 88 (2020) 102828

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

An empirical analysis of dockless bike-sharing utilization and its explanatory factors: Case study from Shanghai, China

Aoyong Li[a], Pengxiang Zhao[b,*], Yizhe Huang[c], Kun Gao[d], Kay W. Axhausen[a]

[a] Institute for Transport Planning and Systems (IVT), ETH Zürich, Zürich CH-8093, Switzerland
[b] Institute of Cartography and Geoinformation, ETH Zürich, Zürich CH-8093, Switzerland
[c] Shanghai Jiao Tong University, 800, Dongchuan RD, Shanghai, China
[d] Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg SE-412 96, Sweden

ARTICLE INFO

Keywords:
Dockless bike-sharing
GPS bike origin-destination data
Empirical analysis
Built environment
Social-demographic characteristics
GWR

ABSTRACT

Revealing dockless bike-sharing utilization pattern and its explanatory factors are essential for urban planners and operators to improve the utilization and turnover of public bikes. This study explores the dockless bike-sharing utilization pattern from the perspective of bike using GPS-based bike origin-destination data collected in Shanghai, China. In this paper, utilization patterns are captured by decoupling several spatially cohesive regions with intensive bike use via non-negative matrix factorization. We then measure the utilization efficiency of bikes within each sub-region by calculating Time to booking (ToB) for each bike and explore how the built environment and social-demographic characteristics influence the bike-sharing utilization with ordinary least squares (OLS) regression and geographically weighted regression (GWR) models. The matrix factorization results indicate that the shared bikes mainly serve a certain area instead of the whole city. In addition, the GWR model shows higher explanatory power (Adjusted R^2 = 0.774) than the OLS regression model (Adjusted R^2 = 0.520), which suggests a close relationship between bike-sharing utilization and the selected explanatory variables. The coefficients of the GWR model reveal the spatial variations of the linkage between bike-sharing utilization and its explanatory factors across the study area. This study can shed light on understanding the demand and supply of shared bikes for rebalancing and provide support for operators to improve the dockless bike-sharing utilization efficiency.

1. Introduction

Bike-sharing, as one of the environmentally friendly and active transport modes, has received notable attention, as it has been shown to be effective in relieving traffic congestion, mitigating transport-related emissions, and improving public health (Wang and Zhou, 2017; Otero et al., 2018; Zhang and Mi, 2018; Barbour et al., 2019). Currently, more than 2000 bike-sharing systems (BSS) are operated in the world due to these advantages,[1] which can be divided into docked and dockless systems. Compared with docked bike-sharing systems that require users to pick up bikes from designated docking stations and drop off bikes at docking stations with available lockers, dockless bike-sharing systems are more flexible, which allow bikes to be picked up and dropped off in any places with available parking space (Mooney et al., 2019). In addition, the fleet size of a dockless bike-sharing system is not constrained by the capacity of docking stations anymore. Dockless systems have evolved as the newest generation of bike-sharing systems due to their convenience, which gained much popularity and were widely accepted (Fishman, 2015; Chen et al., 2020).

Although the adoption and prevalence of bike-sharing systems bring social and environmental benefits, the unbalanced usage of bike-sharing raises concerns. For instance, bikes located in downtown areas of a city may be used frequently, while some other bikes located in the suburbs might not be used at all. This phenomenon is especially prevalent in metropolitan cities owing to inappropriate allocations of bikes and other factors. Additionally, infrequently used bikes will occupy road space and probably mess up cities, especially with dockless bikes. Moreover, with the booming of bike-sharing systems, there is a need to understand what factors influence the unbalanced patterns and their strength. Understanding how those factors impact the usage of bike-sharing could provide policy-making support and reference for determining the allocation of bikes. Considering that dockless bike-

* Corresponding author.
E-mail address: pezhao@ethz.ch (P. Zhao).
[1] http://www.bikesharingworld.com

https://doi.org/10.1016/j.jtrangeo.2020.102828
Received 9 March 2020; Received in revised form 31 July 2020; Accepted 6 August 2020
0966-6923/ © 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).

sharing systems provide more freedom with users to choose origins and destinations of their travels, the utilization patterns and influence factors could be different from those of docked bike-sharing systems. Besides, little attention has been paid to the utilization of dockless bike-sharing systems compared with docked bike-sharing systems. Therefore, it is necessary to recognize those dockless bike-sharing usage pattern and their explanatory factors, which can help to adopt measures for improving the performance and usage efficiency of existing dockless bike-sharing systems.

In recent years, the explosive growth of GPS datasets from bike-sharing systems has stimulated a number of studies in transportation and urban planning, such as on bike-sharing ridership (Rixey, 2013; Wang and Akar, 2019; Yang et al., 2020), bike-sharing demand prediction (El-Assi et al., 2017; Lin et al., 2018; Hua et al., 2020), facility planning for bike-sharing services (Özceylan et al., 2017; Mete et al., 2018; Zhang et al., 2019b), bike accessibility (Li et al., 2020), bike-sharing usage patterns (Faghih-Imani et al., 2017; Guo et al., 2017; Barbour et al., 2019; Xu et al., 2019). Although several studies have explored dockless bike-sharing usages with GPS-based bike data, little attention has been paid to dockless bike-sharing utilization from the perspective of bikes. Previous studies focus on describing bike-sharing usage from the perspective of users, such as bike-sharing ridership, trip demand, riding duration or distance, arrival and departure rates, usage frequency and so on. The knowledge of whether bikes are utilized efficiently and what factors influence bikes' utilization efficiency is limited.

In this paper, the bike-sharing data collected in Shanghai, China is used to investigate the dockless bike-sharing utilization and its influence factors. Our specific research questions are as follows: (1) How do the bike-sharing utilization patterns change in different urban areas? Theoretically, all the shared bikes should be utilized across the whole city. (2) How is the bike-sharing utilization influenced by built environment and social-demographic factors? To solve these problems, we conduct data processing to identify the start and end locations of the trips for each bike, from which the service areas of all bikes are detected with the non-negative matrix factorization method. Second, an indicator called stop duration (SDR) is proposed to quantify the utilization of bike-sharing at the sub-region level. Finally, ordinary least squares regression (OLS) and geographically weighted regression (GWR) models are utilized to examine the relationships between the bike-sharing utilization and the selected variables.

The paper is organized as follows. Section 2 presents a literature review on bike-sharing patterns and bike-sharing analysis. Section 3 introduces the study area and data of this study. Section 4 elaborates the methodology, including bike service areas detection, geographically weighted regression and the explanatory variables. Section 5 presents the results. Discussions and conclusions are provided in Section 6.

2. Related work

In this section, we review some of the relevant research on bike-sharing usage and its explanatory factors. A review of the literature shows that bike-sharing usage studies can be divided into two main categories based on the usage of data sources. The first type contains bike-sharing system state data via a web crawler or other means to capture related information from the websites. For instance, Hampshire and Marla (2012) collected bike-sharing usage data by capturing bike-sharing system state snapshot data via the websites. A panel regression model was used to understand the factors influencing bike-sharing usage. The results indicated that the number of bike stations, population density, and labor market size are important factors. The study by Faghih-Imani et al. (2014) investigated how meteorological data, temporal characteristics, bike infrastructure, land use, and built environment attributes affect bike-sharing usage flows using minute-by-minute bike availability data for all stations in service. They reported that restaurants, commercial enterprises, and universities nearby a station have a significant influence on the arrival and departure rates of the station. Wang et al. (2016) examined the effects of nearby businesses and jobs on trips to and from stations using the number of trips at the station level. It is revealed that the factors, including neighborhood socio-demographics, proximity to the central business district, proximity to other bike-sharing stations, are associated with bike-sharing usage activity. Faghih-Imani et al. (2017) conducted an empirical analysis of bike-sharing usage and rebalancing with station-level occupancy snapshot data from two cities, which are transformed into customer arrivals and departures at the station level. The influence of bike infrastructure, socio-demographic characteristics, and land-use characteristics on bike-sharing usage was explored by developing a mixed linear model. The results showed that bike infrastructure attributes and land-use characteristics are of importance for bike usage. Caulfield et al. (2017) examined the usage patterns of a bike-sharing scheme with the data from a medium-sized city. The results show that short and frequent trips account for the majority of the trips in the scheme. Yang et al. (2019) analyzed the impact of metro service on the dockless bike-sharing patterns from dockless bike availability data with geostatistical analysis and graph-based approach. The results are beneficial to understanding how the new metro service boosted nearby bike demand, thereby supporting spatially detailed urban and transport planning.

In recent years, with the emergence of the new dockless bike-sharing programs that integrate mobile payment and GPS tracking techniques, it is convenient to record the locations and moving trajectories of bikes. These bike trajectory data allow us to analyze bike-sharing usage via spatio-temporal distribution of bikes and gradually become more prevalent. For instance, Shen et al. (2018) uncovered the spatio-temporal patterns of bike usage with the GPS data of dockless bikes in Singapore. Especially, the impacts of bike fleet size, surrounding built environment, access to public transportation, bike infrastructure, and weather conditions were examined using spatial autoregressive models. The findings showed that high land use mixtures, easy access to public transportation, more supportive cycling facilities, and free-ride promotions are positively related to the usage of dockless bikes. Du et al. (2019) developed a model framework to explore the spatio-temporal usage patterns of free-floating shared bikes using massive GPS data produced by the bike-sharing system. They found that residential area, park and green area, and population size have the largest influence on the usage frequency. Zhang et al. (2019a) proposed the concept of biking islands that are defined as geographical areas of interest with a high concentration of bike usage, and a framework to identify them with bike trajectory data. This study can help understand the travel behavior of cyclists and the urban structures used for cycling. The study by Ji et al. (2020) performed a comparison of usage regularity and its determinants between docked and dockless bike-sharing systems using the smart card data of a docked bike-sharing system and GPS trajectory data of a dockless bike-sharing system in Nanjing, China. By analyzing the influence of travel characteristics and built environment factors on the regularity of bike-sharing usage, the results provide insights on improving the service of bike-sharing systems to operators and governments.

There are also several studies identifying the factors affecting bike-sharing usage with questionnaire data. For example, Bachand-Marleau et al. (2012) examined the factors influencing the likelihood of using shared bike systems and frequency of use by conducting a survey in Montreal, Canada. The proximity of home to docking stations was found to have the most substantial effect on the likelihood for use of shared bike systems. Guo et al. (2017) explored the factors influencing bike-sharing usage among the bike-sharing user population by collecting a questionnaire survey data in Ningbo, China. A bivariate ordered probit model was developed to examine the associations between bike-sharing usage and the selected factors. It is found that gender, household bike ownership, travel time, bike-sharing station location have a significant influence on bike-sharing usage. Barbour et al. (2019)

conducted a statistical analysis of bike-sharing usage by collecting data on the bike-sharing usage of registered bike-sharing users with a web-based survey. Gender, age, income, household size, commute type and length, and vehicle ownership have a substantial influence on bike-sharing usage.

Despite this large body of literature on bike-sharing usage, most of the aforementioned studies, however, are concentrated on the bike-sharing usage from the perspective of bike users. This study intends to investigate dockless bike-sharing utilization and its explanatory factors from the perspective of bikes. In particular, time to booking is used to explore bike-sharing utilization patterns in this study. Moreover, GWR models are employed to examine the spatial non-stationarity of the relationship between bike-sharing utilization and the influence factors.

3. Study area and data description

Shanghai is one of the largest cities in China and the world, with a population of 24.24 million and an area of 6341 km^2. Shanghai's GDP is 3.268 trillion Chinese Yuan in 2018, accounting for 6.6% of China's total GDP. As one of the largest bike-sharing markets in the world, Shanghai has 1.5 million dockless bikes in July 2017. In this paper, the whole city is divided into sub-regions (SR) according to neighborhoods, which are also the basic units of the sixth nationwide population census. We select the sub-regions that are within or intersected by the outer ring road as the study area, containing 133 sub-regions. The study area totally or partly contains 11 districts, including Pudong, Baoshan, Huangpu, Jiading, Xuhui, Putuo, Minhang, Yangpu, Changning, Hongkou, Jing'an districts. Huangpu River, the largest river in Shanghai, divides the study area into “Pudong” (the east side of Huangpu River) and “Puxi” (the west side of Huangpu River), as shown in Fig. 1. The inset at the right corner of Fig. 1 presents the whole city of Shanghai.

The bike-sharing data used in this study were provided by Mobike,[2] which is one of the biggest dockless bike-sharing companies. This dataset spans the period from August 26^{th} to September 8^{th} in 2018, covering 14 days. Each record contains basic trip information, including trip ID, start time, start location (longitude and latitude), end time and, end location. Considering that some trips exceed the study area, we filter out those trips whose start locations or end locations are outside the study area.

After removing the trips outside the study area, statistical analysis of basic bike usage patterns is conducted, as shown in Fig. 2. Fig. 2a and b illustrate the distributions of trip distance and trip duration, respectively. It can be observed that most of the trips are less than 5 km or 1 h. Concretely, the proportions of the trips with a distance less than 5 km and the ones with duration less than 1 h are 96% and 98.3%, respectively, which demonstrates that bike is a short distance transport mode. Fig. 2c shows the temporal patterns of bike usage frequency on workdays and weekends. The weekday pattern displays two peak periods (8:00–9:00 and 18:00–19:00), which may reflect the typical commuting pattern on weekdays. The significant peak periods, however, fail to appear on the weekend, which indicates that people's biking activities distribute relatively uniform during the daytime on the weekend. Furthermore, Fig. 2d describes the distribution of the number of trips per bike, which shows that about 20% of bikes are rodden less than ten times. Based on Fig. 2, the trips with an abnormal length or duration are also excluded. To avoid the influence of the bikes with few trips, we only consider bikes that have at least 14 trips (i.e., at least one trip per day). Finally, 19.4 million trips from 348,037 bikes are included.

Another three datasets, Point of Interest (POI), the smart card data, and population data, are used to calculate the explanatory factors. POI data is collected from AutoNavi[3] (also known as Gaode Map, a Chinese navigation company). Each POI record contains name, address, category, and location (longitude and latitude). According to the category information, this dataset contains 17 types of POIs, which are shown in Table 1. The smart card data are collected from the Shanghai Public Transport Card Company. The data consists of basic subway trip information, including start time and subway station ID, end time, and subway station ID. The population data is sourced from the Sixth National Census of China. The data provides information on the resident population and the density of population at neighborhood level (i.e., the sub-region in this study).

4. Methodology

4.1. Detecting service areas of bikes

In order to explore the bike-sharing utilization pattern, we detect the service areas of bikes with non-negative matrix factorization (NMF). First, the start locations of the trips are first mapped to the corresponding sub-regions (SR). A visitation sequence is then defined as the sequence of consecutive start locations based on SR, namely $\mathscr{S}$: $\{SR^{t1}, SR^{t2}, \ldots, SR^{tn}\}$, for each bike. Furthermore, we obtain a visitation frequency matrix $V_{|M|\times|N|}$, where M represents the number of bikes, N is the number of sub-regions. The elements denote the visitation frequency of each sub-region for each bike.

Non-negative matrix factorization (NMF) is applied to the visitation frequency matrix V. NMF is a robust technique to detect hidden patterns in a matrix, which decouples a non-negative matrix into two non-negative matrices of lower rank. The two matrices represent the meta information of bikes and sub-regions in terms of bike visitation frequency. In addition, NMF has the inherent property to cluster these bikes and sub-regions. Mathematically, given a non-negative matrix $V_{|M|\times|N|}$ and a pre-defined rank K, the NMF algorithm decouples V into two smaller non-negative matrix W and H, i.e., $V_{|M|\times|N|} \approx W_{|M|\times K} \times H_{K\times|N|}$, and targets to minimize the difference between the left-hand side and the right side. W is the coefficient matrix and H represents the dictionary matrix. Theoretically, for any rank K (meeting $K \ll min\,[|M|, |N|]$), the NMF algorithm can be performed and generates a result. The key issue is to determine the optimal factor rank K. Several mature rank selection criteria, such as cophenetic correlation coefficient (Brunet et al., 2004), dispersion (Kim and Park, 2007) and the residual sum of squares (RSS) curve (Hutchins et al., 2008), have been developed. In this study, we adopt the cophenetic correlation coefficient criterion, which is widely used to select the optimal K (Kang and Qin, 2016). The value of k is selected where the magnitude of the cophenetic correlation coefficient begins to fall (Brunet et al., 2004).

4.2. Measuring utilization efficiency of bikes

Time to booking (ToB) for a bike represents the duration that the bike is booked again after the previous trip has ended (Guidon et al., 2019), which is employed to measure the utilization efficiency of bikes. Compared with the existing bike-sharing usage indicators (e.g. riding duration, usage frequency, and turnover rate), ToB is capable of better reflecting supply and demand of a certain area, thereby measuring the utilization efficiency of bikes in the area. Taking riding duration as an example, it depends on the origin and destination of trip and cyclist's route choice, which is unrepresentative of demand and supply of one region. Theoretically, if one sub-region is suffering from oversupply of bikes, some bikes may not be used any more for a long time after they are parked here. Otherwise, if bikes within the region are in short supply, the bikes may be used again for a short time. In addition, although usage frequency and turnover rate can evaluate bike-sharing utilization efficiency to some extent, it measures bike-sharing usage from the perspective of users and is partly influenced by ride duration and uses' behavior. For example, given two bikes A and B during a

[2] https://mobike.com/
[3] https://lbs.amap.com/

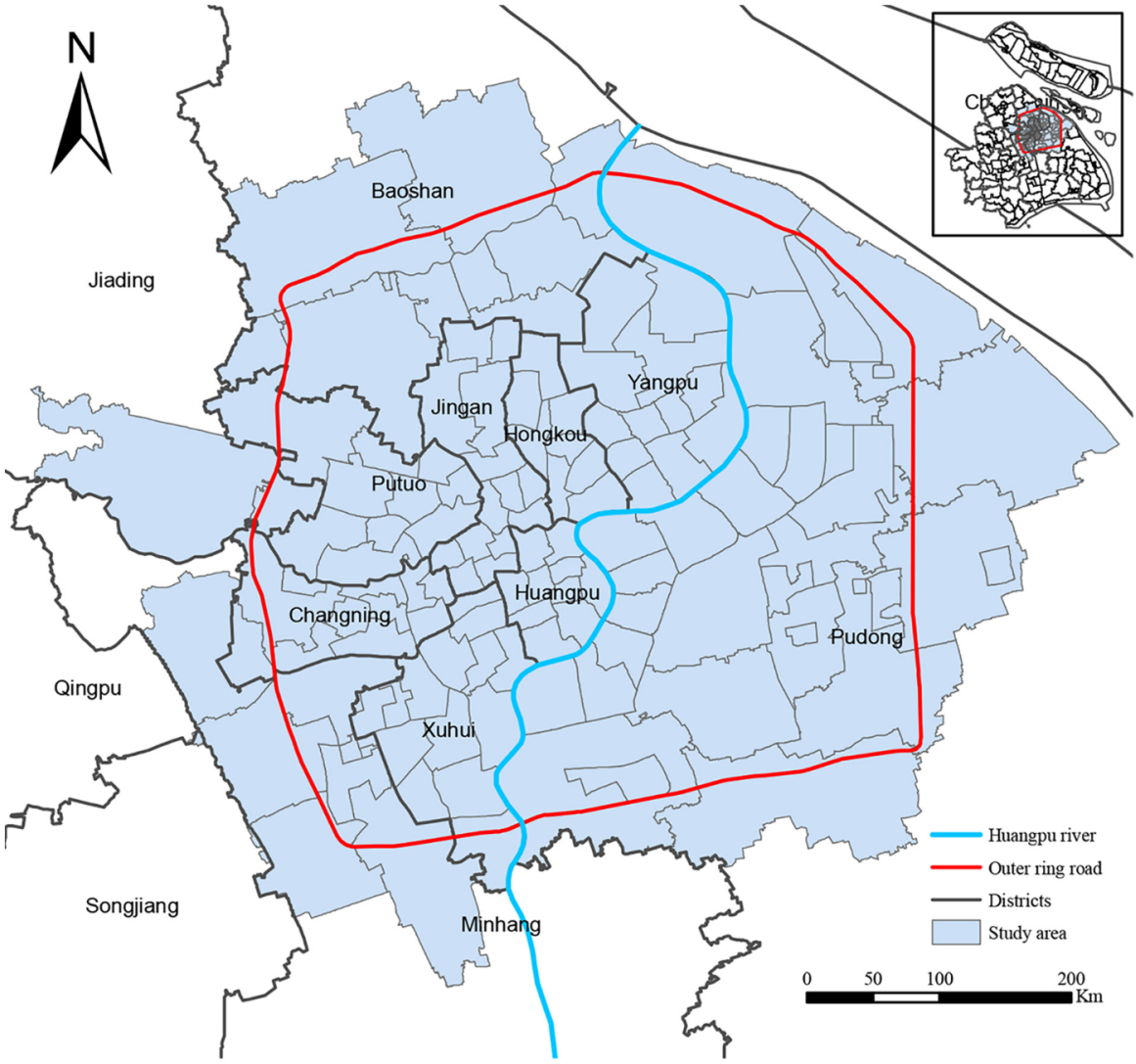


Fig. 1. The study area.

certain period, it is supposed that bike A was rode for a long time during one trip, while bike B was rode for several short trips during the same period. It can be seen that A has lower usage frequency and turnover rate even if it was used more efficiently during this period. Hence, ToB is superior in measuring bike-sharing utilization frequency in one area by examining the relationship between supply and demand perfectly.

To calculate the time to booking for each bike, we first extract all pairs of consecutive trips for each bike. Given a trip of one bike, its consecutive previous trip cannot always be obtained due to various reasons (e.g., data quality). We only consider the trips whose start points are less than δ meters away from the end point of the previous trip. The δ is determined based on two aspects, 1) the accuracy of the positioning device adopted by bike-sharing system,[4] 2) the distribution of distances between two consecutive trips in the data (Fig. 3), which shows most distances are not too long. Thus, the "inflection point" $\delta = 50$ is adopted as the threshold to calculate the time to booking:

$$ST_{m,t} = S_{m,t} - D_{m,t-1} \quad (1)$$

where $S_{m,t}$ is the start time of the current trip for bike m; $D_{m,t-1}$ is the end time of the previous trip for the same bike.

[4] https://kknews.cc/tech/q285v38.html

Based on Eq. (1), we obtain the ToB values for all the included bikes during the whole period. Fig. 4 shows the distribution of ToB, which displays a long-tail distribution. Concretely, the mean, median, and standard deviation are 175.1, 29.6, and 452.0 min respectively. The ToB for one bike reflects the bike's utilization efficiency, which is an important indicator for operators. To analyze the utilization pattern of the bike-sharing system and how it is affected by the influence factors in urban space, we further project ToB values of bikes to the corresponding sub-regions. To characterize the bike-sharing utilization at the sub-region level, we define the median of all the ToB values within each sub-region as stop duration (SDR), representing the indicator of the corresponding sub-region. In a sub-region, higher SDR represents a lower utilization efficiency of bikes, and a lower SDR represents a higher utilization efficiency of bikes.

Fig. 5 shows the statistical and spatial distribution of SDR per sub-region. The SDR varies between 12.3 and 156.8 min with the mean at 38.1 min (and a median of 29.8 min). From Fig. 5a, it is reported that bikes in most sub-regions are used again quickly (less than 1.5 h), and only a few sub-regions have a high SDR. In addition, the spatial distribution shows a clear spatial pattern across the study area. Sub-regions near the center have a lower SDR, increasing as one moves away from the center. Especially, these sub-regions with low SDR are

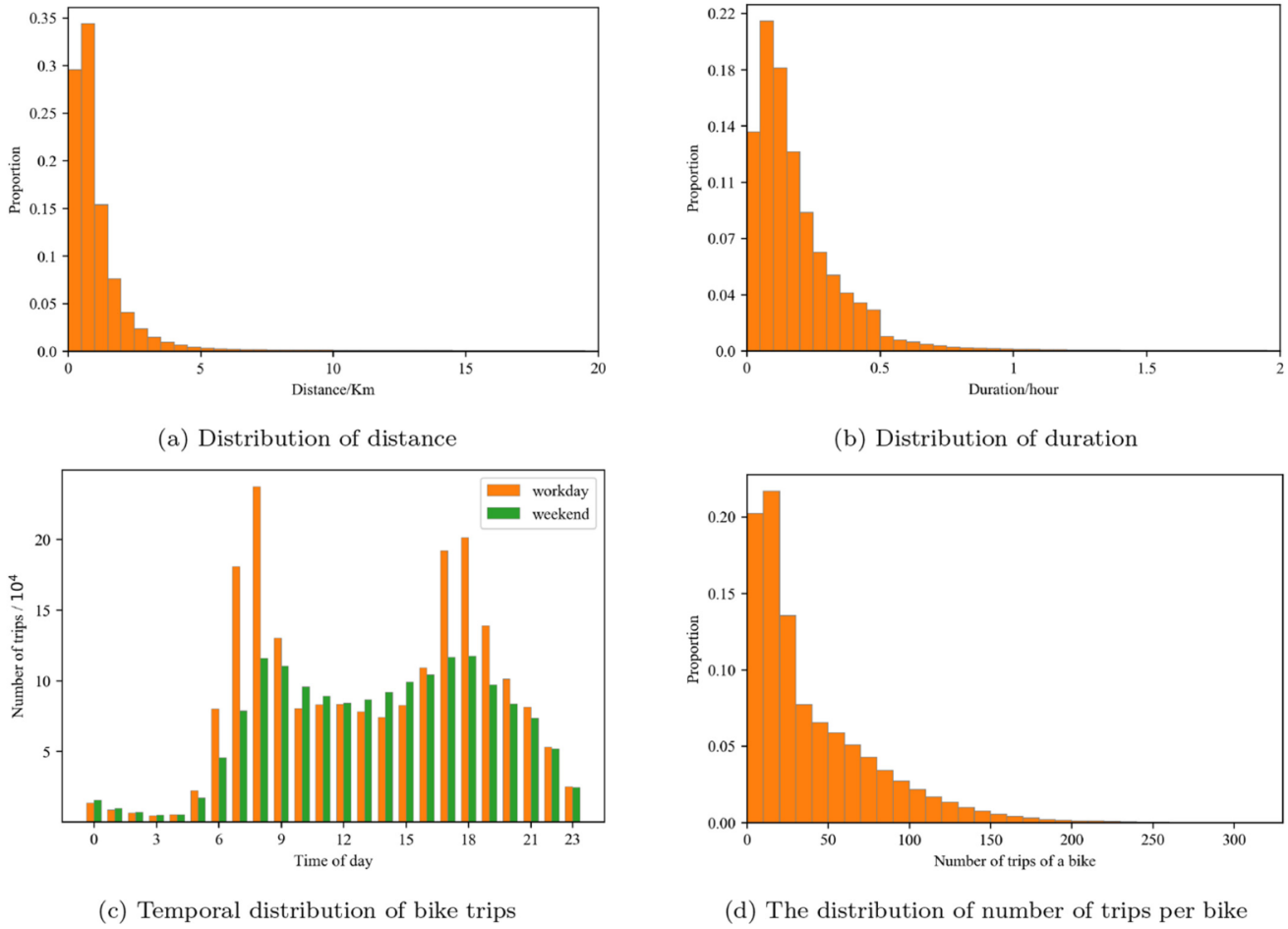


Fig. 2. Statistical analysis of basic bike usage patterns.

Table 1
The POI Categories.

Code	POI Category	Code	POI Category
1	Car Service	2	Car Sales
3	Car Repair	4	Motorcycle Service
5	Restaurant	6	Shopping
7	Daily Life	8	Sports
9	Hospital Related	10	Hotel Related
11	Tourist Attraction	12	Residence
13	Governmental Organization	14	Education Related
15	Transport related	16	Finance Service
17	Enterprises		

concentrated within the inner ring road, which are the most developed areas of Shanghai. The sub-regions with the highest SDR are distributed outside of the outer ring road. Furthermore, we also measure the spatial autocorrelation of SDR based on Moran's I by using ArcGIS. The Moran's I is 0.32 ($z - score = 6.615$ and $p < 0.001$), which indicates SDR values are correlated in space.

4.3. Exploring influence factors of utilization efficiency

To investigate the spatial heterogeneity of SDR in urban space, we employ a Geographically Weighted Regression (GWR) model to quantify the relationship between the stop duration of each sub-region and the selected explanatory factors, e.g. socio-demographic features, POI characteristics, and traffic facilities' convenience. In this section, we briefly introduce the GWR model and the corresponding independent variables.

4.3.1. Geographically weighted regression

Generally, the ordinary least squares regression (OLS) is used to explore the global relationship between dependent and independent variables, which assumes independent variables have the same effect on the dependent variable across the study area. The OLS regression model is denoted by Eq.(2).

$$Y_i = \alpha_0 + \sum_k \alpha_k X_{ki} + \varepsilon_i \tag{2}$$

where Y_i represents the dependent variable, α_k is the estimated coefficient of the independent variable X_{ki}, and ε_i is the residual error.

The estimated regression coefficients obtained from OLS regression models are constant for all sub-regions of the entire study area due to the global characteristics of this model. However, the SDRs are correlated with each other in space (see Section 4.2). The GWR model is a local spatial statistical method, which assumes the nearby observations have greater a influence on the variable than the distant observations (Brunsdon et al., 1996). The model quantifies the spatial effect by conducting a separate OLS regression for each location. The local regression models only incorporate the explanatory variables and dependent variables of locations falling within the bandwidth of target locations. Hence, GWR considers spatial autocorrelation and spatial heterogeneity of observations, which has been widely used to examine spatial non-stationarity of regression models (Yang et al., 2017; Soltani et al., 2018; Zhao et al., 2020). The GWR model is formulated as:

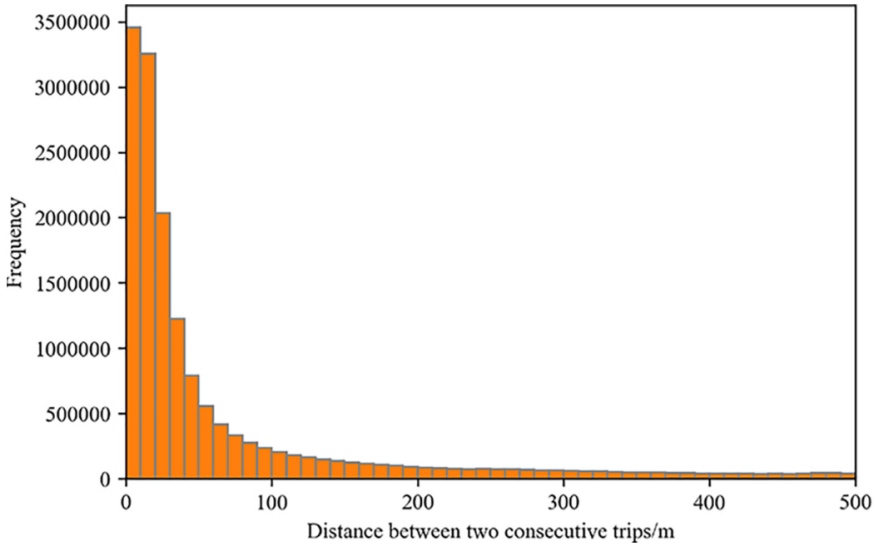


Fig. 3. The distribution of distance between two consecutive trips. Most of values are short distance.

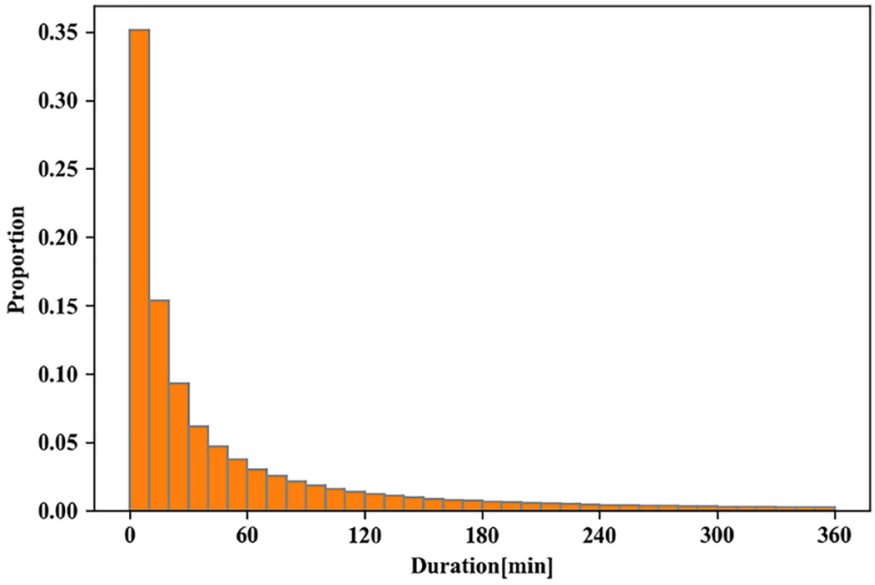


Fig. 4. The distribution of ToB.

$$Y_i = \alpha_0(u_i, v_i) + \sum_k \alpha_k(u_i, v_i) X_{ki} + \varepsilon_i \quad (3)$$

where (u_i, v_i) represents the coordinate of location i and $\alpha_k(u_i, v_i)$ is the regression coefficient of independent variable X_{ki} at location i. Weighted least squares are used to estimate the coefficients of the GWR model, where the weights capture the spatial effect of local observations on the observation at location i. The spatial weight matrix is calculated by the Gaussian kernel function, which models the spatial effects of surrounding observations by Gaussian distance decay within the bandwidth. Bandwidth has a great impact on the estimations of coefficients. It is determined by the mgwr package in python.[5]

4.3.2. Explanatory variables

In this section, the candidate explanatory variables are derived from the built environment and social-demographic characteristics, which have been demonstrated to be closely associated with human mobility behavior.

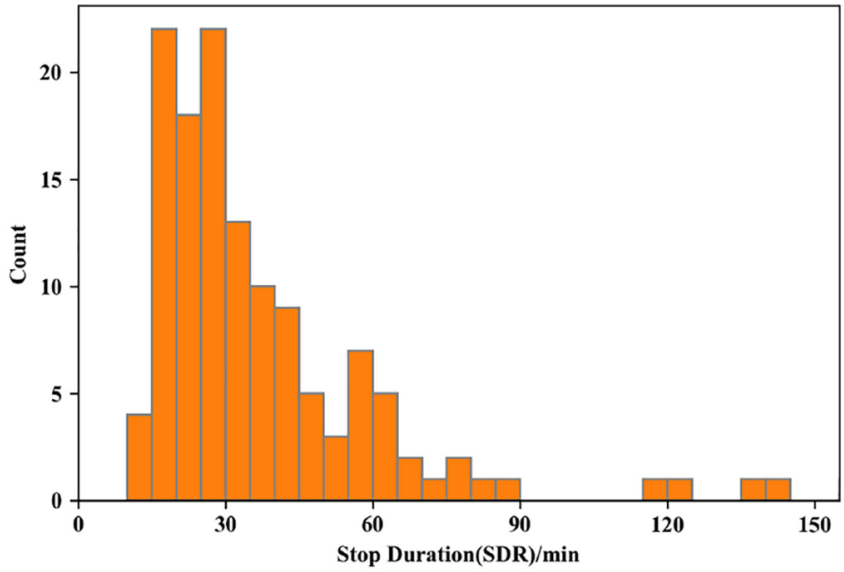
Land use has a significant influence on human mobility and further influences the utilization efficiency of bike-sharing. Due to the advances in sensor and positioning technologies, the percentages of 17 types of POI (See Table 1) for each sub-region are calculated to indicate the land use characteristics of each sub-region. In addition, the mixture of POI is calculated to represent the degree of the land use mix. Entropy is employed to measure the mixture of POI in each sub-region, which is formulated as

$$pm_i = -\sum p_{i,k} \log(p_{i,k}) \quad (4)$$

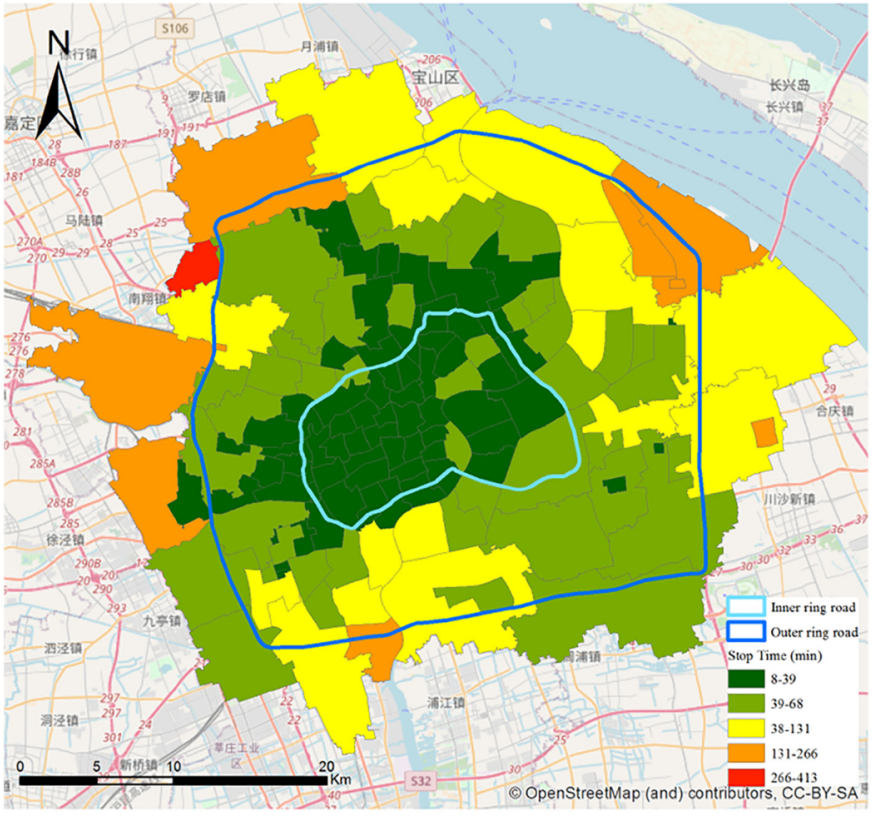
where pm_i refers to the POI mix of the ith sub-region; $p_{i,k}$ denotes the proportion of kth POI in ith sub-region.

To evaluate the degree of transport convenience, bus stop density (BSD), subway station density (SSD), and distance from sub-region to the nearest subway station (D2NSS) have been widely used in previous

[5] https://github.com/pysal/mgwr



(a) Statistical distribution



(b) Spatial distribution

Fig. 5. The statistical and spatial distribution of SDR per sub-region.

studies. BSD and SSD are defined as the number of bus stops and subway stations within each sub-region divided by the area of the sub-region respectively, as illustrated in Fig. 6a. D2NSS is defined as the distance between the center of sub-region and its nearest subway station. One limitation of the three above mentioned variables is that they regard transit stations or sub-regions as points. However, the influence area of either transit station or sub-region is a polygon. For example, although the values of BSD and SSD for the central sub-region in Fig. 6a are 0, the central sub-region is definitely influenced by the transition stations from adjacent sub-regions. In terms of D2NSS, it is inaccurate to represent a sub-region by the center point of it, especially when the sub-region is large. To solve this problem, we propose a novel indicator

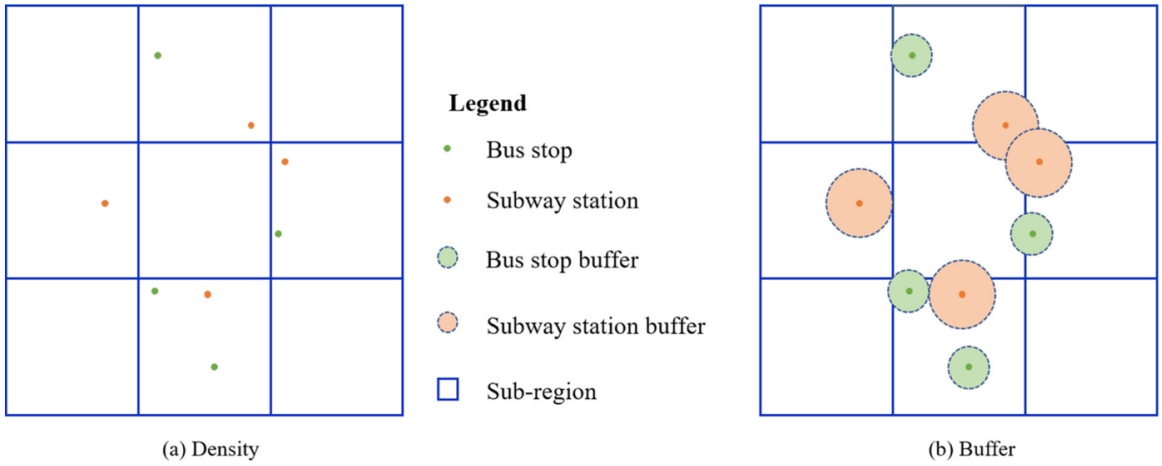


Fig. 6. Sketch map on transport convenience related factors.

to measure transport convenience, namely the proportion of overlapping area between influence area of transit station and the corresponding sub-region (See Fig. 6b). The influence area of each transit station is defined as a buffer with a radius of x meters. Hence, how to determine the optimal value of x is noteworthy. In this study, all distances between every two nearest bus stops or subway stations are calculated. x is determined as the mean of these distances. Thus, $x = 158$ and $x = 1025$ are set as the radii of influence areas for bus stops and subway stations, respectively. The two variables are denoted as ARBus and ARSubway, which can be calculated based on the overlapping area between the buffer of bus stop or subway station and the corresponding sub-region.

Moreover, we incorporate two socio-demographic variables, namely density of permanent resident population (DRP) and the average flow of subway stations (AFSS). Intuitively, sub-region with more residents should have higher travel demand and short stop duration. Average flow of subway stations represents the average number of passengers leaving and going into the nearest n subways each hour, which is formulated as:

$$Avg\_flow = \frac{1}{m}\sum_{i}^{m}\frac{(HI_i + HO_i)}{2} \tag{5}$$

where HI_i is the average number of passengers going into the subway station each hour; HO_i is the average number of passengers leaving the subway station each hour; m denotes the nearest m subway stations (we adopt $m = 3$ here), which is selected according to the Euclidean distance from the center of each sub-region to the subway stations.

Overall, 22 independent variables, including 18 POI related variables (the percentages of 17 types of POI and the mixture of POI), two transport related variables, and two population related variables, are explored, which share different units and ranges. To make them comparable, they are therefore scaled using the following equation:

$$x_{sij} = \frac{x_{ij} - min(X_i)}{max(X_i) - min(X_i)} \tag{6}$$

where x_{sij} is the scaled value for variable i at point j; x_{ij} is the raw value of x_{sij}; X_i is the variable i; and $min(X_i)$ and $max(X_i)$ are the minimum and maximum of variable X_i, respectively.

5. Results and discussion

5.1. Spatial characteristics of bikes utilization patterns

In this subsection, we explore the spatial characteristics of bikes' utilization patterns based on the non-negative matrix factorization (See subsection 4.1). The trips for each bike are extracted to construct visitation sequence, thereby generating the visitation frequency matrix. To determine the optimal parameter K, we test different values (i.e. from 2 to 12) to select the best rank according to the cophenetic correlation coefficient, as shown in Fig. 7. The cophenetic coefficient does not always decrease along increasing rank K. $K =$ 2, 6, and 11, which start to decrease, have better decomposition results compared with other ranks. In addition, all the sub-regions are divided into K groups, which are displayed in Fig. 8. The first column illustrates the consensus map, which indicates the probability that a pair of sub-regions are assigned to the same cluster. Each consensus map presents K separating blocks. The second column expounds the spatial distribution of each cluster corresponding to one block in the consensus map. It can be seen that the sub-regions in the same cluster are adjacent with each other in space, which indicates that the bikes are mainly used in a limited area instead of the whole study area. Notably, the clusters are becoming smaller with the increment of K values. For instance, from $K = 2$ to $K = 6$, cluster 0 in Fig. 8(a) is divided into cluster 1 and 3 in Fig. 8(b), while cluster 1 is separated into 0, 2, 5, and 4. However, their spatial distributions still show some similar patterns. For example, the Huangpu river is a natural barrier that divides the sub-regions of either side into different clusters no matter the K, as the bikes are not allowed to cross the river via the bridge or tunnel.

Intuitively, the sub-regions within each block have various visitation frequencies, which display the bikes' service patterns. These patterns are represented as a set of eigenbehaviors (EB), which are derived from the matrix W. For the sake of illustrative ease, these patterns are mapped on the urban space, as shown in Fig. 9. Without loss of generality, the figure only presents the eigenbehaviors for Rank = 6. Since the size of sub-regions varies widely, instead of the frequency of pick-ups, we map the density of visitation frequency (i.e., the number of pick-ups per square kilometer). In Fig. 9, all the sub-figures share the same range of the visitation density across the six scenarios, and the deeper color represents the high density of pick-ups.

Obviously, the density of pick-ups for each pattern demonstrates a remarkable heterogeneity. In each sub-figure, we can easily find the sub-region with a higher value, which indicates pick-ups are mainly concentrated in these areas. In addition, the densities between the six patterns are also different. EB1 (Fig. 9a), EB2 (Fig. 9b), and EB6 (Fig. 9f) have a high value, while it is not easy to find a deeper color in EB3 (Fig. 9c), EB4 (Fig. 9d), and EB5 (Fig. 9e).

In Fig. 9a, the deep color is mainly concentrated in Wujiaochang, Siping Road Street, etc. The sub-region with the highest value is Wujiaochang, which is one of the sub-centers and central business districts in Shanghai. With its development, Wujiaochang is also a transport

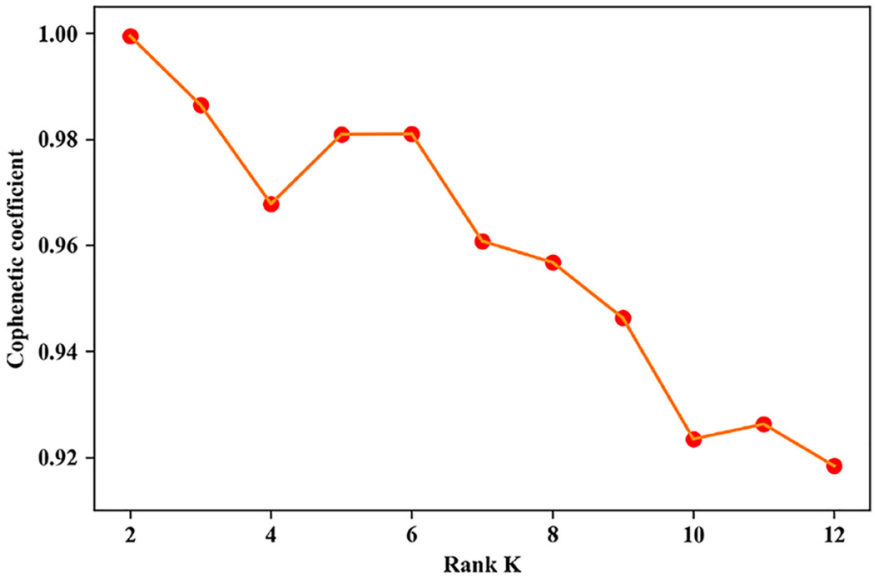


Fig. 7. Rank Selection.

hub. In addition, Fudan university is also located in this sub-region. These facts also make Wujiaochang produce the majority of the trips in this block. For Siping Road Street sub-region, a campus of Tongji University and Shanghai Yangpu High School are located there. Together with several residential blocks, Siping Road Street is also a sub-region with a high density of bike trips. For EB2 (Fig. 9b), the visitations are mainly distributed in Pudong district, such as Lujiazui sub-region with the highest density. Lujiazui contains central business district (CBD) and several residential blocks. The CBD is a national-level financial district with many famous views such as Oriental Pearl Tower and Shanghai Tower. Although bikes are not allowed on most roads of this CBD, Lujiazui sub-region still displays a high visitation frequency according to the data. The sub-regions with high visitation frequency of EB6 (Fig. 9f) are mainly located in Huangpu, Hongkou, Jing'an, Putuo, Changning districts, which are the center of Shanghai. For example, the sub-regions with the highest density are mainly distributed within the inner ring road, like Jingansi street, Xujiahui street, West Nanjing Road street, which are the most developed area in Shanghai. With regards to the other three patterns, the sub-regions with middle density are mainly distributed outside of the inner ring road. Compared with EB1, EB2, and EB6, fewer bike trips are generated in these areas. EB3 (Fig. 9c) is distributed in the northeast of the study area. Within this EB, the sub-regions with middle density are also distributed in Jing'an and Putuo district. EB4 (Fig. 9d) reports a bike-sharing usage pattern in Puxi. Although it has some overlaps with EB2, its highest-value sub-regions are distributed between the inner ring road and the outer ring road. As to EB5 (Fig. 9e), it mainly describes the usage pattern in the southeast of the study area. The sub-regions with the highest value are adjacent to the boundary of the study area, which reveals that the bikes used in these sub-regions could also be connected with the regions outside the study area. In addition, these highest-value sub-regions are the center of the Minhang district.

5.2. Relationship between utilization pattern and impact factors

In this section, we explore the relationship between the SDR of sub-regions and selected factors by using OLS and GWR models.

5.2.1. OLS regression analysis results

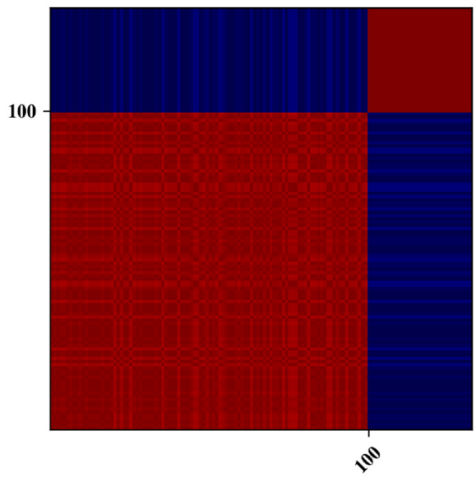
We first employ the OLS regression model to examine the effect of the variables on the SDR. In our models, transport related variables and population related variables are always included. However, it is infeasible to use all of the POI related variables because some of them are highly related to each other. Therefore, we attempt different categories to build OLS models. Before applying the model, variable inflation factors (VIF) is used to test multicollinearity for each group of the independent variables. Ultimately, the optimal OLS regression model is selected according to the corrected Akaike information criterion (AICc) and the significance of independent variables. The determined explanatory variables and their descriptions are shown in Table 2.

Table 3 displays the result of the selected OLS regression model. The adjusted R^2 is 0.520, which indicates that the selected independent variables can explain 52.0% of the variations. In this model, both ARSubway and ARBus have negative impacts on the stop duration. It means that if the influence area of subway stations or bus stops occupies a high percentage of sub-region, the bikes within that sub-region will be used again after a short stop time. In addition, ARSubway is significant while ARBus does not have a significant association with SDR, which indicates that the subway stations, compared with bus stops, have more remarkable impact on bike utilization. With regard to the two population related variables, namely DRP and AFSS, both of them have a negative influence. This is because the sub-regions with high DRP and AFSS normally have larger demand for bike-sharing. For the POI related variables, POI mixture is also negatively related to SDR, which indicates the sub-regions with various types of POI require more sharing bikes than the sub-region with single type of POI. In the selected model, we consider the percentages of four types of POI, including restaurants, daily life service, residence, and commercial facilities, which represent four types of activities. The coefficients of PRt and PRc are negative, while PDLS and PCF have a positive influence. In addition, both PRc and PDLS are significant at the 0.05 level, but residence and commercial facilities do not show significant influence.

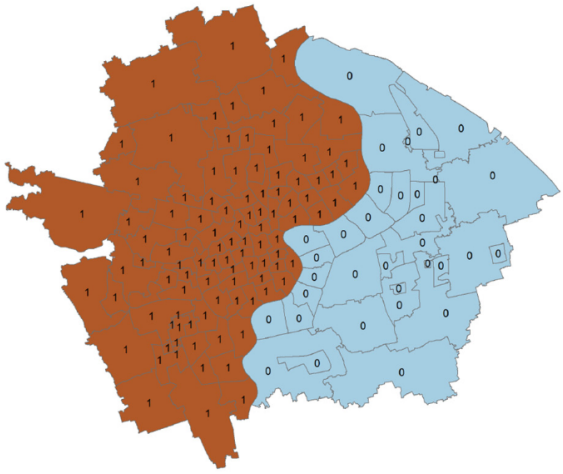
Although the OLS regression model is capable of providing a global understanding of the relationship between SDR and the explanatory variables, spatial non-stationary inspection is neglected. Therefore, the relationship is further explored with GWR model.

5.2.2. GWR analysis results

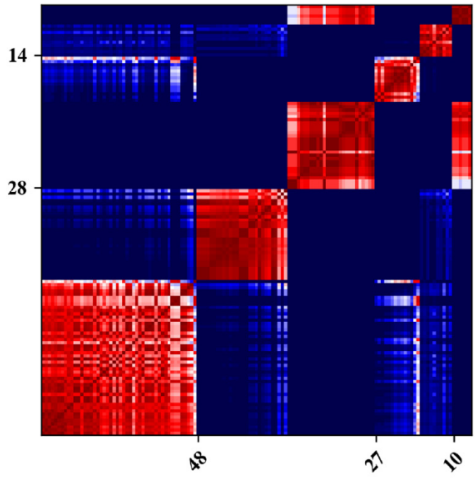
In this section, we further examine the relationships between SDR of sub-regions and the explanatory variables based on GWR model. The model yields a better result with adjusted $R^2 = 0.774$, which is higher than that of OLS regression model (i.e. adjusted $R^2 = 0.52$). In addition,



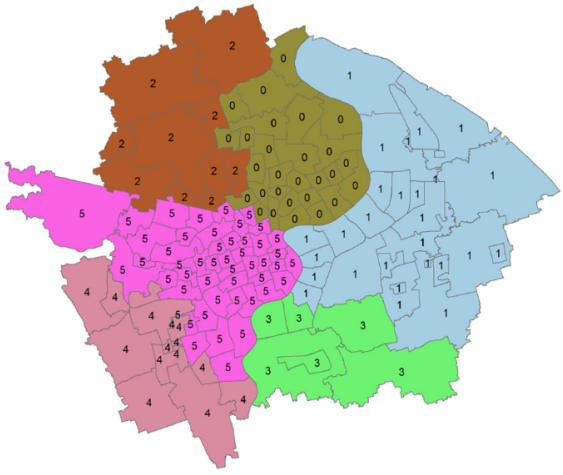
Rank = 2



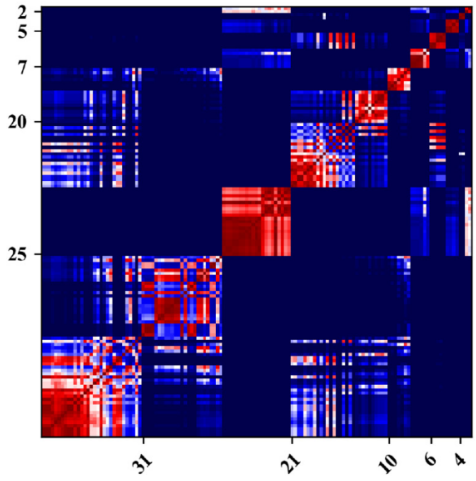
(a) Rank = 2



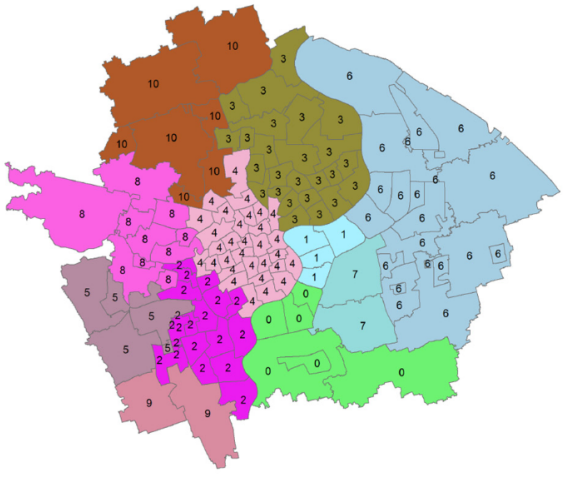
Rank = 6



(b) Rank = 6



Rank = 11



(c) Rank = 11

Fig. 8. Consensus map and cluster map for rank = {2, 6, 11}. The first column shows consensus maps and the second column shows the cluster map in space. The sub-figures in the same column are derived from the same Rank. For consensus maps, the number in x-axis and y-axis represents the number of sub-regions in each

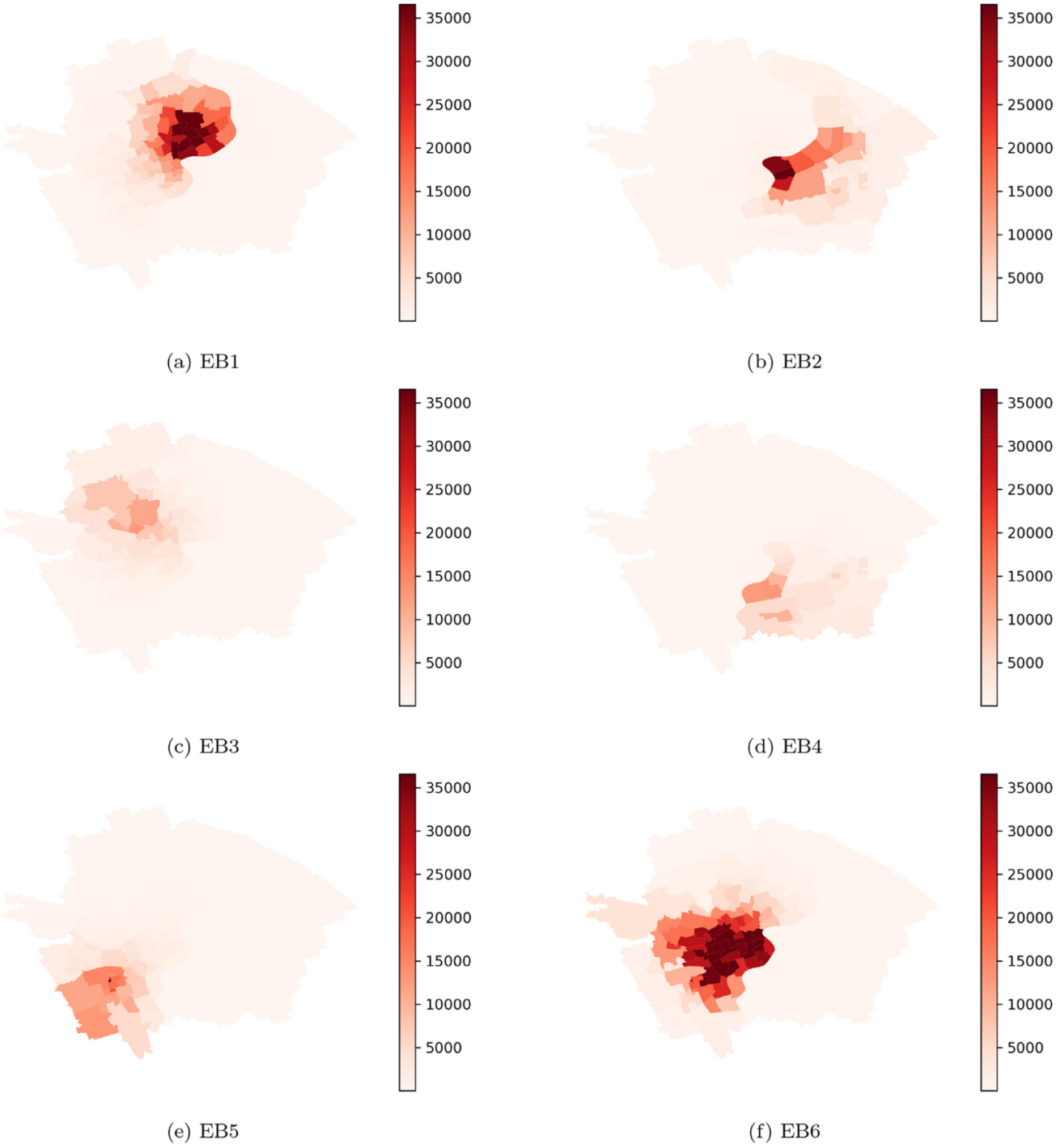


Fig. 9. Spatial patterns of eigenbehaviors for Rank = 6. The value is the density of visitation. The deeper the color is, the higher the value is.

the Akaike Information Criterion (AIC) of GWR model (i.e. 1095.4) is lower than that of OLS regression model (i.e., 1158.1). It can be concluded that the GWR model presents better explanatory power than the OLS regression model for exploring the influence of explanatory factors on bike-sharing utilization. Moreover, we also examine the spatial distribution of the standard residuals from the GWR model, as shown in Fig. 10. It can be observed that only 5 local regression models (3.7%) fail the residual tests, which correspond to the blue and red sub-regions. The spatial autocorrelation of standard residuals is further examined by calculating the global Moran's I. The Moran's I (−0.019) result indicates that the standard residuals from the GWR model display random distribution at the 5% significance level.

Table 4 shows the estimated results of GWR model, which lists the descriptive statistics of the varying regression coefficients, including Mean, standard deviation (Std.), Minimum, lower quantile (25%), Median, upper quantile (75%), and Maximum. The descriptive statistical indexes provide a general understanding of the distribution characteristics of regression coefficients. For example, POIMix and PRc (percentage of residence) display large standard deviation values, which partly indicates that the distribution of their coefficients is more

Table 2
The explanatory variables for OLS and GWR.

Variable name	Description
ARSubway	The ratio between the influence area of subway stations and the area of sub-region
ARBus	The ratio between the influence area of bus stops and the area of sub-region
DRP	The density of permanent resident population
AFSS	Average flow of subway stations
POIMix	Mixture of POI
PRt	Percentage of Restaurants
PDLS	Percentage of Daily life services (e.g. travel agency)
PRc	Percentage of Residence
PCF	Percentage of commercial facilities

Table 3
The result of the OLS regression model.

Independent variable	Coef.	Std.	t-Statistic	p
ARSubway	−42.552	7.224	−5.890	0.000***
ARBus	−5.225	13.408	−0.390	0.697
DRP	−19.588	10.140	−1.932	0.056*
AFSS	−18.834	8.740	−2.155	0.033**
POIMix	−13.915	23.104	−0.602	0.548
PRt	−52.730	18.948	−2.783	0.006***
PDLS	34.010	16.926	2.009	0.047**
PRc	−13.843	21.009	−0.659	0.511
PCF	10.029	14.230	0.705	0.482
Constant	97.662	27.520	3.549	0.001***
R^2	0.552			
Adjust R^2	0.520			
AICc	1158.1			
No. Observations	133			

*** Represents the significance of the regression coefficient at the 0.01 level.
** Represents the significance of the regression coefficient at the 0.05 level.
* Represents the significance of the regression coefficient at the 0.1 level

dispersed than other variables. For the variables like PDLS (percentage of daily life services), more than 50% of sub-regions have a positive impact on stop duration, whereas average flow in all sub-regions has a negative effect on stop duration. Next, we investigated the spatial variations of these explanatory variables on stop duration by projecting these coefficients on the sub-regions.

Fig. 11 shows the spatial distribution of estimated coefficients for all the selected variables from GWR model. The red color indicates that the corresponding explanatory variables have a positive influence, and blue color represents a negative influence. The deeper the color is, the greater the influence is. The GWR model is able to quantify the spatial varying influence of each independent variable across study area. The spatial characteristics of the influence of each explanatory variable are analyzed.

For the two transport related variables, namely ARSubway (Fig. 11a) and ARBus (Fig. 11b), which display a negative influence on stop duration in most sub-regions. It means that if there are more subway stations or bus stops around one sub-region, the bikes within the sub-region will be used again after a short stop period with higher probability. This confirms the previous findings that the bikes within the urban sub-regions near public transportation stations (e.g., subway station, MRT station, bus stop) normally display high bike usage pattern (Shen et al., 2018; Xu et al., 2019). It can be concluded that bike-sharing might serve a number of last-mile trips. Note that positive associations are identified in previous studies compared with the negative association in this study, which can be attributed to the measurement of bike-sharing utilization. For instance, the indicators like the arrival and departure rates, usage frequency, the average number of trips made per bike are employed to measure the bike-sharing usage (Hampshire and Marla, 2012; Faghih-Imani et al., 2017; Shen et al., 2018; Du et al., 2019), higher values of which correspond to more bike-sharing usage. On the contrary, as mentioned in Section 4.2, a smaller stop duration for bikes represents higher bike-sharing utilization in this work. In addition, as shown in Fig. 11a, the sub-region with lower influence of ARSubway are mainly concentrated on the Bund, Yu Garden, Nanjing road, which are the most developed areas in Shanghai and have highly dense subway stations. One possible reason is that people can reach the

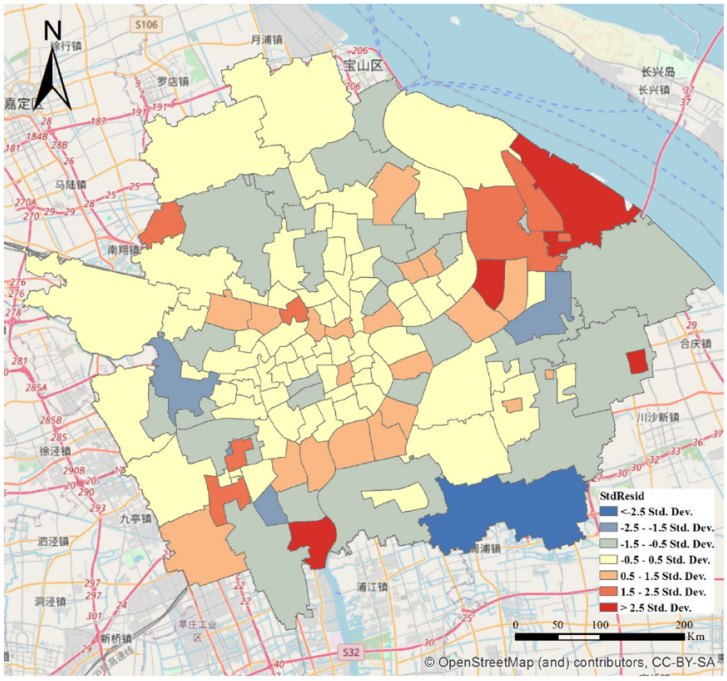


Fig. 10. Spatial distribution of standard residuals of GWR model.

Table 4
The results of GWR model.

Independent variable	Mean	Std.	Minimum	25%	Median	75%	Maximum
ARSubway	−29.8	20.8	−66.7	−49.3	−25.0	−12.7	0.5
ARBus	−7.1	14.6	−45.8	−13.3	−6.5	3.2	17.0
DRP	−15.0	12.6	−41.9	−20.0	−14.2	−8.8	16.0
AFSS	−15.3	11.2	−37.1	−21.3	−12.0	−6.7	2.2
POIMix	−23.5	43.3	−166.2	−27.6	−20.2	−0.8	73.2
PRt	−44.2	38.9	−149.7	−56.6	−27.2	−18.4	9.4
PDLS	8.6	24.2	−27.6	−9.0	−2.4	22.9	79.2
PRc	−8.5	34.1	−74.9	−33.0	−6.9	22.9	39.9
PCF	−9.0	21.2	−49.0	−23.0	−9.8	0.0	50.6
R^2	0.829						
Adjusted R^2	0.774						
AICc	1095.40						
No. Observations	133						

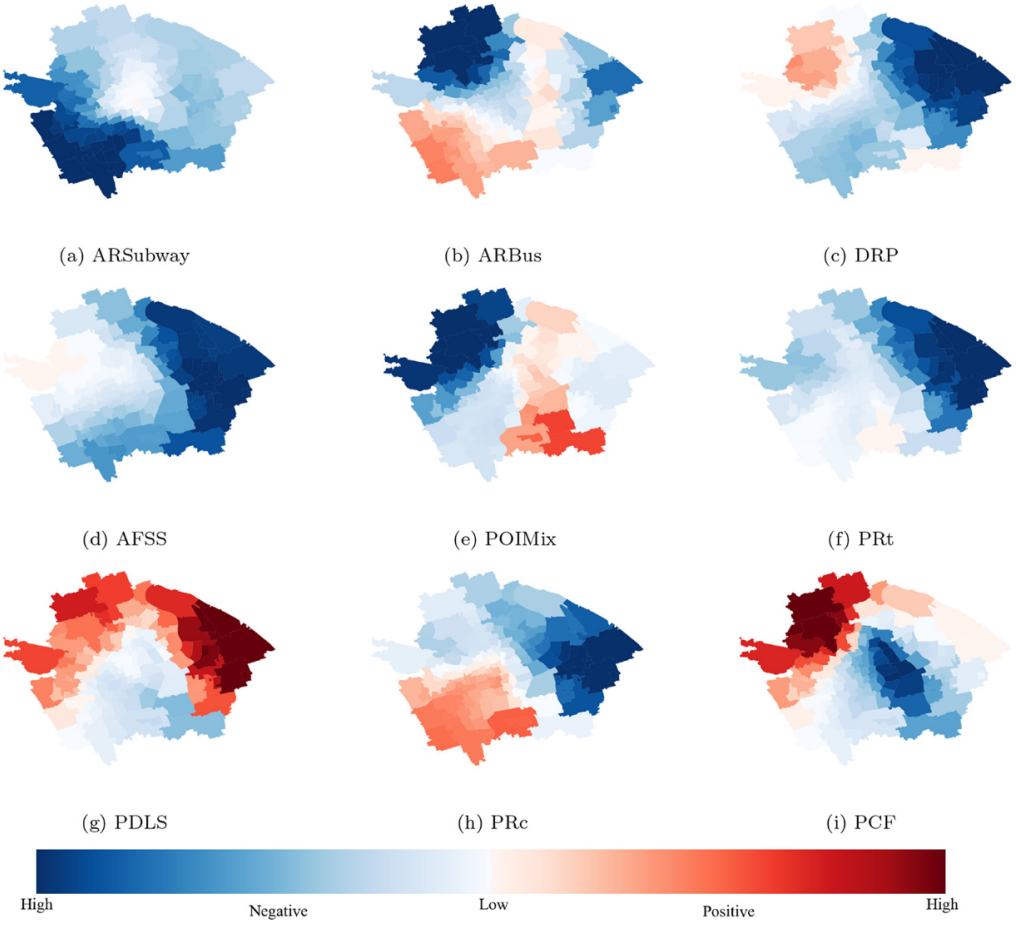


Fig. 11. Spatial distribution of varying regression coefficients.

subway stations by walking easily in these sub-regions. Furthermore, cycling is not allowed in some roads within these areas. For ARBus, the sub-region with large negative influence are mainly concentrated in Baoshan district, and the sub-regions with positive coefficients are mainly distributed in Minhang and Pudong district. Although the positive influence is existed in some sub-regions, they only display small influence. It can be attributed to high bus stop density in these sub-regions and ease of accessibility for bus stops.

In terms of the two population related variables, they mainly show a negative influence on SDR. For population density (DRP) (Fig. 11c), most of the sub-regions present negative associations with stop duration SDR. We can conclude that most of the sub-regions with a dense population normally present a small SDR (i.e. high bike-sharing utilization). The conclusion is supported by the findings of previous studies that high population density tends to generate and attract more biking trips and then improve bike-sharing usage (Hampshire and Marla, 2012). However, the GWR model also examined the non-stationarity of the variable in this study. It is reported that the sub-regions with strong negative associations are mainly distributed in the periphery of Pudong district, while the sub-regions with small positive coefficients are mainly concentrated in Baoshan and Jiading district. This result occurs may due to the incomplete living facilities in Pudong. In this situation, high population density normally brings high travel volume. In Baoshan and Jiading district, considering that most sub-regions are distant from

metro stations, people depend heavily on the combination of bus and subway. Hence, population density displays a low influence on the bike-sharing utilization in these sub-regions. In Fig. 11d, it is reported that AFSS always has a negative influence on stop duration. Especially, the sub-regions in Pudong district display high impact on SDR. Compared with DRP, the AFSS can motivate the usage of bike-sharing in the whole Pudong district.

Regarding POI mixture (POIMix), the sub-regions with negative coefficients are mainly concentrated on the northwest of study area, namely the sub-regions in Baoshan, and the positive influence is mainly distributed in Pudong (Fig. 11e). It is interesting to note that the increasing mixture of POI can improve the diversity of activities in Puxi, which makes SDR becomes shorter. However, the sub-regions with high POI Mixture in Pudong district do not attract more people to use shared bikes. Although previous studies demonstrated that POI mixture can be beneficial to more diverse activities and attract more riders (Shen et al., 2018; Wang and Akar, 2019), spatial non-stationarity is identified from the result. One possible explanation is that although some sub-regions in Pudong also have high POI mixture like the sub-regions in Puxi, the sub-regions in Pudong could have different attractions on biking trips considering the various POI types and incomplete living facilities in Pudong. From Fig. 11f, we can observe that the percentage of restaurants (PRt) has a negative impact in almost all the sub-regions. It means that bike-sharing system provides a convenient way for residents to reach a restaurant. In addition, it can be seen that the sub-regions in Pudong have a higher influence than that of the sub-regions in Puxi, especially the central part of the study area. A possible explanation is that more restaurants are concentrated in the center of the study area. It is easy to reach them by walking instead of using a bike. Conversely, regarding the percentage of Daily life services (PDLS), it is found that PDLS presents a positive influence on most sub-regions in the periphery of study area. In addition, the higher percentage of residence (PRc) generally means more residents, which generates a higher travel demand. PRc shows a negative influence on SDR in most sub-regions (Fig. 11h), which also conforms to the conclusion that residential POI positively influences the bike-sharing usage (Ji et al., 2020). However, a higher percentage of residence does not motivate to decrease the SDR in several sub-regions in the south of the study area. In terms of the percentage of commercial facilities (PCF), they do not always show a positive influence in the whole study area. As displayed in Fig. 11i, PCF has a negative influence in the southeast of the study area, which is mainly in Pudong district.

In summary, the influence of the selected explanatory variables shows spatial variations across the study area. The OLS regression model examines the global relationship between the variables and stop duration for each sub-region. The GWR model identifies spatial variations of influence across the study area by considering spatial autocorrelation and heterogeneity. For instance, for the variables such as ARBus, POIMix, PRc, and PCF, the proportions of sub-regions with positive and negative coefficients are comparable to each other, which indicates complicated relationships between these variables and SDR of sub-regions. This is why the coefficients of these variables are not significant in the OLS regression model. This knowledge provides an insight into the relationship between bike-sharing usage and built environment as well as socio-demographic characteristics, which can assist decision-making for the management and planning of bike-sharing.

6. Conclusions

Understanding the dockless bike-sharing utilization pattern and its explanatory factors can be beneficial to urban planners and operators to improve the utilization efficiency and turnover of public bikes. However, little attention has been paid to investigating this particular pattern from the perspective of bikes. To bridge the research gap, this study explores the dockless bike-sharing utilization pattern and its influence factors by conducting an empirical study using a GPS-based bike origin-destination data collected in Shanghai, China. The trips' origin-destination of each bike are extracted from the data, which are used to calculate Time to Biking values for each bike. These ToB values are further mapped to the corresponding sub-regions to describe and measure the bike-sharing utilization efficiency at the sub-region level. The main contributions of this work are summarized as follows.

First, we detect the service areas of bikes to explore the bike-sharing utilization pattern with non-negative matrix factorization at the sub-region level. The visualization of matrix factorization results shows that several spatially cohesive regions with intensive bikes exchange are identified. In particular, the bike-sharing utilization patterns in six scenarios are interpreted. It is indicated that the bikes have their own service areas instead of being utilized across the whole study area. The results are beneficial for the operators to rebalance the shared bikes, thereby further improving the utilization efficiency of shared bikes and saving parking spaces.

Second, we define an index called stop duration (SDR) to delineate the bike utilization at the sub-region level, which describes the utilization efficiency of bikes within each sub-region. In addition, the OLS and GWR models are used to quantify the relationships between SDR and the selected explanatory variables related to transportation, land use, and social-demographic factors. Among these explanatory variables, we define three novel features, including the proportion of overlapping area between influence area of transit stations (i.e. subway and bus) and the corresponding sub-region, and the average flow of subway stations. The first and second variables consider the influence of transit as polygons instead of points to represent the convenience of access. The average flow of subway stations, which consider the flow of subway stations as a dynamic population index. The coefficients of the OLS regression models demonstrate that ARSubway, the density of resident population (DRP), average flow of subway stations (AFSS), and the percentage of restaurants (PRt) are negatively related to SDR at a significance level of 0.05, while percentage of daily life services (PDLS) is positively related to SDR at the same significance level. According to the corrected Akaike information criterion (AICc) and adjusted R-square, the GWR model presents the remarkable improvement of effectiveness (Adjusted R^2 = 0.774, $AICc$ = 1095.4) compared with the OLS regression model (Adjusted R^2 = 0.52, $AICc$ = 1158.1). Furthermore, the varying regression coefficients revealed the spatial varying of the influence of various factors on the stop duration, which also explains why several variables are not significant in the OLS regression model.

It is worth noting that this study has several limitations that should be considered in the future. First, this study considers the whole bike data to detect operation patterns and conduct utilization analysis but ignores the influence of time. We should put much effort to understand the bikes' operation pattern in specific periods, like busy traffic periods and free flow periods. Second, all of our findings are in an aggregate level, namely population census zone in our study. However, the spatial scale is an important concept in geography, which may result in different conclusions. It still needs to further expand the work on different spatial scales. Finally, SDR, which reflects the interaction of demand and supply, shows the performance of measuring the utilization efficiency of bike-sharing. It would be interesting to study how stop duration can be used to measure bike-sharing providing a good service.

Author agreement statement

We declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that all of us have approved the order of authors listed in the manuscript.

We understand that the Corresponding Author is the sole contact for

the Editorial process. He is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs.

Author contributions

Aoyong Li: Conceptualization, Methodology, Investigation, Writing - original draft, Writing - review &editing.

Pengxiang Zhao: Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing.

Yizhe Huang: Resources, Writing - review & editing.

Kun Gao: Writing - review & editing.

Kay W. Axhausen: Writing - review & editing, Supervision.

Acknowledgements

The research was supported by Chinese Scholarship Council Scholarship. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsors.

References

Bachand-Marleau, J., Lee, B.H., El-Geneidy, A.M., 2012. Better understanding of factors influencing likelihood of using shared bicycle systems and frequency of use. Transp. Res. Rec. 2314 (1), 66–71.

Barbour, N., Zhang, Y., Mannering, F., 2019. A statistical analysis of bike sharing usage and its potential as an auto-trip substitute. J. Transp. Health 12, 253–262.

Brunet, J.-P., Tamayo, P., Golub, T.R., Mesirov, J.P., 2004. Metagenes and molecular pattern discovery using matrix factorization. Proc. Natl. Acad. Sci. 101 (12), 4164.

Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. Geogr. Anal. 28 (4), 281–298.

Caulfield, B., O'Mahony, M., Brazil, W., Weldon, P., 2017. Examining usage patterns of a bike-sharing scheme in a medium sized city. Transp. Res. A Policy Pract. 100, 152–161.

Chen, Z., van Lierop, D., Ettema, D., 2020. Dockless bike-sharing systems: what are the implications? Transp. Rev. 40 (3), 333–353 ISSN 0144-1647 1464-5327.

Du, Y., Deng, F., Liao, F., 2019. A model framework for discovering the spatio-temporal usage patterns of public free-floating bike-sharing system. Transp. Res. C. 103, 39–55.

El-Assi, W., Mahmoud, M.S., Habib, K.N., 2017. Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. Transportation 44 (3), 589–613.

Faghih-Imani, A., Eluru, N., El-Geneidy, A.M., Rabbat, M., Haq, U., 2014. How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (bixi) in Montreal. J. Transp. Geogr. 41, 306–314.

Faghih-Imani, A., Hampshire, R., Marla, L., Eluru, N., 2017. An empirical analysis of bike sharing usage and rebalancing: evidence from Barcelona and Seville. Transp. Res. A Policy Pract. 97, 177–191.

Fishman, E., 2015. Bikeshare: a review of recent literature. Transp. Rev. 36 (1), 92–113 ISSN 0144-1647 1464–5327.

Guidon, S., Becker, H., Axhausen, K.W., 2019. Avoiding stranded bicycles in free-floating bicycle-sharing systems: Using survival analysis to derive operational rules for rebalancing, paper presented at the IEEE Intelligent Transportation Systems Conference (ITSC 2019).

Guo, Y., Zhou, J., Wu, Y., Li, Z., 2017. Identifying the factors affecting bike-sharing usage and degree of satisfaction in Ningbo, China. PLoS One 12 (9).

Hampshire, R.C., Marla, L., 2012. An Analysis of Bike Sharing Usage: Explaining Trip Generation and Attraction from Observed Demand, Paper Presented at the 91st Annual meeting of the transportation research board, Washington, DC. pp. 12–2099.

Hua, M., Chen, X., Zheng, S., Cheng, L., Chen, J., 2020. Estimating the parking demand of free-floating bike sharing: a journey-data-based study of Nanjing, China. J. Clean. Prod. 244, 118764.

Hutchins, L.N., Murphy, S.M., Singh, P., Graber, J.H., 2008. Position-dependent motif characterization using non-negative matrix factorization. Bioinformatics 24 (23), 2684–2690 (ISSN 1367-4803).

Ji, Y., Ma, X., He, M., Jin, Y., Yuan, Y., 2020. Comparison of usage regularity and its determinants between docked and dockless bike-sharing systems: a case study in Nanjing, China. J. Clean. Prod. 120110.

Kang, C., Qin, K., 2016. Understanding operation behaviors of taxicabs in cities by matrix factorization. Comput. Environ. Urban. Syst. 60, 79–88 (ISSN 0198-9715).

Kim, H., Park, H., 2007. Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. Bioinformatics 23 (12), 1495–1502 (ISSN 1367-4803).

Li, A., Huang, Y., Axhausen, K.W., 2020. An approach to imputing destination activities for inclusion in measures of bicycle accessibility. J. Transp. Geogr. 82, 102566.

Lin, L., He, Z., Peeta, S., 2018. Predicting station-level hourly demand in a large-scale bike-sharing network: a graph convolutional neural network approach. Transp. Res. C. 97, 258–276.

Mete, S., Cil, Z.A., Özceylan, E., 2018. Location and coverage analysis of bike-sharing stations in university campus. Bus. Syst. Res. J. 9 (2), 80–95.

Mooney, S.J., Hosford, K., Howe, B., Yan, A., Winters, M., Bassok, A., Hirsch, J.A., 2019. Freedom from the station: spatial equity in access to dockless bike share. J. Transp. Geogr. 74, 91–96.

Otero, I., Nieuwenhuijsen, M., Rojas-Rueda, D., 2018. Health impacts of bike sharing systems in europe. Environ. Int. 115, 387–394.

Özceylan, E., Mete, S., Çil, Z.A., 2017. Optimizing the location-allocation problem of bike sharing stations: A case study in gaziantep university campus, paper presented at the Proceedings of 14th International Symposium on Operations Research of Slovenia. pp. 141–146.

Rixey, R.A., 2013. Station-level forecasting of bikesharing ridership: station network effects in three us systems. Transp. Res. Rec. 2387 (1), 46–55.

Shen, Y., Zhang, X., Zhao, J., 2018. Understanding the usage of dockless bike sharing in Singapore. Int. J. Sustain. Transp. 12 (9), 686–700.

Soltani, A., Pojani, D., Askari, S., Masoumi, H.E., 2018. Socio-demographic and built environment determinants of car use among older adults in Iran. J. Transp. Geogr. 68, 109–117.

Wang, K., Akar, G., 2019. Gender gap generators for bike share ridership: evidence from citi bike system in New York city. J. Transp. Geogr. 76, 1–9.

Wang, M., Zhou, X., 2017. Bike-sharing systems and congestion: evidence from us cities. J. Transp. Geogr. 65, 147–154.

Wang, X., Lindsey, G., Schoner, J.E., Harrison, A., 2016. Modeling bike share station activity: effects of nearby businesses and jobs on trips to and from stations. J. Urban Plan. Dev. 142 (1), 04015001.

Xu, Y., Chen, D., Zhang, X., Tu, W., Chen, Y., Shen, Y., Ratti, C., 2019. Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system. Comput. Environ. Urban. Syst. 75, 184–203.

Yang, H., Lu, X., Cherry, C., Liu, X., Li, Y., 2017. Spatial variations in active mode trip volume at intersections: a local analysis utilizing geographically weighted regression. J. Transp. Geogr. 64, 184–194.

Yang, Y., Heppenstall, A., Turner, A., Comber, A., 2019. A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile. Comput. Environ. Urban. Syst. 77, 101361.

Yang, H., Zhang, Y., Zhong, L., Zhang, X., Ling, Z., 2020. Exploring spatial variation of bike sharing trip production and attraction: a study based on chicago's divvy system. Appl. Geogr. 115, 102130.

Zhang, Y., Mi, Z., 2018. Environmental benefits of bike sharing: a big data-based analysis. Appl. Energy 220, 296–301.

Zhang, Y., Lin, D., Liu, X.C., 2019a. Biking islands in cities: an analysis combining bike trajectory and percolation theory. J. Transp. Geogr. 80, 102497.

Zhang, Y., Lin, D., Mi, Z., 2019b. Electric fence planning for dockless bike-sharing services. J. Clean. Prod. 206, 383–393.

Zhao, P., Xu, Y., Liu, X., Kwan, M.-P., 2020. Space-time dynamics of cab drivers' stay behaviors and their relationships with built environment characteristics. Cities 101, 102689.